INNOVATION

The Allen Brain Atlas: 5 years and beyond

Allan R. Jones, Caroline C. Overly and Susan M. Sunkin

Abstract | The Allen Brain Atlas, a Web-based, genome-wide atlas of gene expression in the adult mouse brain, was an experiment on a massive scale. The development of the atlas faced a combination of great technical challenges and a non-traditional open research model, and it encountered many hurdles on the path to completion and community adoption. Having overcome these challenges, it is now a fundamental tool for neuroscientists worldwide and has set the stage for the creation of other similar open resources. Nevertheless, there are many untapped opportunities for exploration.

The Allen Brain Atlas (ABA) is a Web-based, three-dimensional atlas of gene expression throughout the adult mouse brain, comprising a genome-wide image database of *in situ* hybridization (ISH) data, a high-resolution anatomical reference atlas and a suite of integrated search, navigation and visualization tools. From its inception, the ABA was intended to provide the scientific community with a powerful resource that would have a broad, positive impact on neuroscience research. At that time, the human and mouse genomes had been sequenced. With full inventories of available genes, the next challenge was to uncover their biological functions, and knowing where in the brain genes are expressed was expected to provide important clues to both gene and brain functions. In addition, technologies for high-throughput data production, management and informatics were maturing, making genome-wide studies and the integration of genomic and neuroanatomical data feasible.

This article looks back on the 5 years from the inception of the ABA to the present, highlighting some of the challenges that were faced in executing the project and the contributions that it has made to neuroscience. We discuss the advantages and caveats of using this unique resource, discuss how it is currently being used and point to untapped opportunities for further exploration. Finally, we describe the ever-expanding suite of related resources that have become available since the ABA was launched, and comment on those that will be coming in the next few years.

Development of the atlas

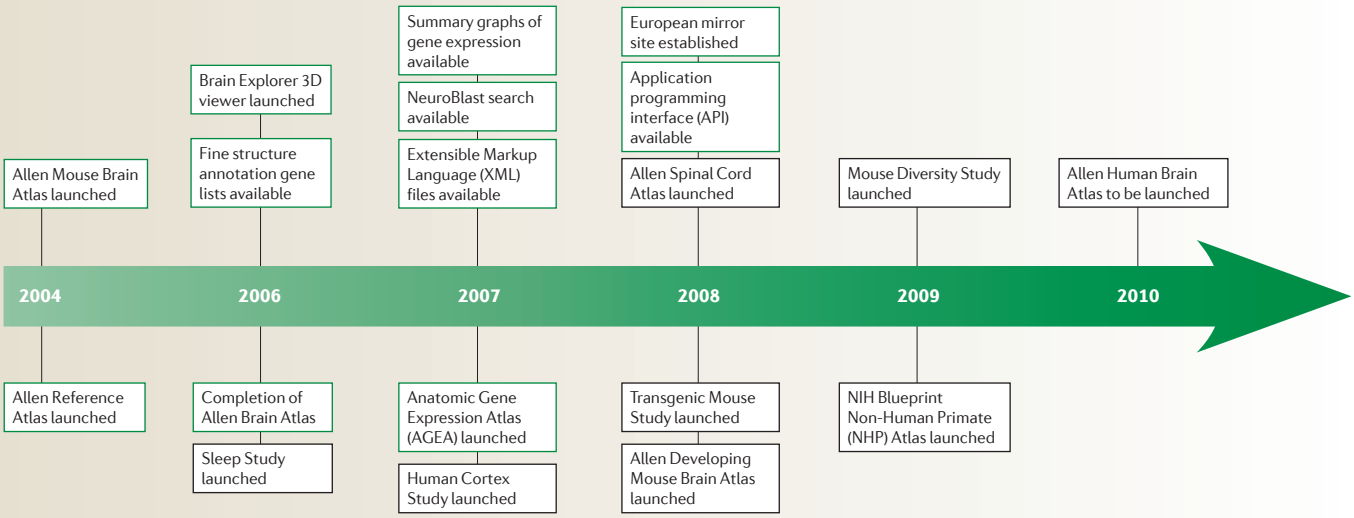
The ABA has its roots in a series of brainstorming sessions that began in 2001 and were led by James Watson, Steven Pinker and others. In these sessions, Paul Allen gathered together groups of scientists with interests ranging from molecular biology to human neuropsychology and asked "What can be done to help propel neuroscience research forward?" David Anderson of the California Institute of Technology put forward the concept of the ABA during these early discussions.

In January 2002, the US National Institutes of Health (NIH)held a meeting to chart the course of neuroscience research in the post-genomic era[1]. There, a host of scientists organized by Marc Tessier-Lavigne and Lubert Stryer concluded that "enormous benefit will derive from a systematic, large-scale, and organized effort to generate a molecular brain map for humans and the mouse"[1]. At that time, as part of the Gene Expression Nervous System Atlas (GENSAT) project[2–4], the National Institute of Neurological Disorders and Stroke (NINDS) was funding two complementary approaches to map gene expression in the mouse brain: one based on creating bacterial artificial chromosome (BAC)-transgenic mouse reporter lines for individual genes and one developed by Gregor Eichele and colleagues at the Max Planck Institute in Hannover, Germany, and implemented by Eichele and Christina Thaller at the Baylor College of Medicine, Texas, USA, using colorimetric ISH to map gene expression[5]. Soon thereafter, the NIH channelled its funding towards the transgenic mouse effort, which has subsequently generated over 800 transgenic reporter mouse lines, most of which have been deposited in the Mutant Mouse Regional Resource Centers (MMRRC)[6].

Several participants at the NIH meeting were part of the ABA advisory council, which was chartered in the autumn of 2002 and chaired by Marc Tessier-Lavigne. After considerable debate over various techniques, including microarray approaches, it was decided that the nascent Allen Institute for Brain Science would use the ISH techniques developed by the Baylor group to pursue a comprehensive, genome-wide effort to map gene expression throughout the mouse brain. The timeline was aggressive: they aimed to survey over 20,000 transcripts and make the data publicly available online in 3 years. The ABA team quickly formed to scale up the laboratory processes to achieve the throughput required. To meet ambitious early data production goals, Eichele and Thaller were contracted to perform the work for ~2,000 genes at Baylor while the process was integrated, automated and scaled up at the fledgling Allen Institute facility.

The initial public release of the ABA in December 2004 offered only a glimpse of what it would later become. Rather than waiting for a complete data set, the ABA data were rolled out in stages to put them in the hands of the research community as early as possible. Following a trajectory typical of software releases, the ABA was expanded, enhanced and improved over time. Even after the entire genome-wide data set was completed and released in September 2006, improvements continued. Data were refined, accessibility was

Timeline | **The Allen Brain Atlas and related public resources**



Boxes outlined in green relate to the original Allen Brain Atlas; those outlined in black relate to other public resources.

increased and innovative tools were added in order to further expand the openness and utility to the research community (TIMELINE).

Early challenges

The first data release comprised brain-wide image series for approximately 2,000 genes — 10% of the final data set — and offered basic search, visualization and other tools for accessing and mining the data. This was an important and successful milestone, demonstrating that ISH data generation could be industrialized and that a Web-based application could be developed to enable easy public access to the resulting image data (a sizable technological challenge beyond that faced for genomics databases). However, the true test of its acceptance, usefulness and impact in neuroscience was to come. A new concept for the community, the ABA was not immediately accepted. As with the Human Genome Project[7,8], the ABA project found sceptics across the neuroscience community early on.

Accessibility. Early in the project, there were active discussions within the Allen Institute community about data accessibility. In particular, there was debate about whether or not future atlas maintenance and development efforts could or should be sustained through cost recovery from commercial users or from intellectual property reach-through to downstream inventions emanating from atlas use. In the end, those options were dismissed, and it was determined that making the ABA available to everyone for free would have the greatest impact on neuroscience research.

The degree of accessibility was also considered. To get the full breadth of the data set out to the community quickly, software engineering resources were first focused on organizing and making the data viewable online, with features for deeper data access coming later. As a consequence, some early users were unhappy that the full-resolution ISH images and informatics data were not available for download. At that time, the images could be requested by e-mail through the website. Several requests were fulfilled for full-resolution digital images, but few users desired them because of the extremely large file size (~20–40 Mb per brain section), the file format (JPEG2000) and the difficulty in processing these files on a basic desktop computer. Now, the ABA offers a mechanism for directly downloading high-resolution images in the more universal JPEG file format. Similarly, the related informatics data were also available by request and online accessibility was enabled with the launch of the application programming interface (API), discussed below.

What's the catch? Some early users were wary. Although from the beginning access to the ABA did not require users to register or submit personal information (a standard that is more relaxed than the common practices of other resources), it did involve a click-through end-user licensing agreement (EULA). The terms were basic, aiming to allow open use of the data for research while protecting the Allen Institute's ownership. Nevertheless, the EULA was misunderstood by some and, despite revisions for clarification, caused some academic scientists concern. Ironically, pharmaceutical companies, arguably those with the most to lose from any intellectual property restrictions, accepted the EULA and were early and large consumers of the data. Ultimately, the EULA was removed, although users are still governed by the basic terms of use (available on the atlas website).

Caveats for users. Despite its wealth of data, the ABA is not without its limitations. Importantly, the ABA comprises primary data, the interpretation and relevance of which must be determined by the end user. Issues with the platform and caveats of using the ABA have been documented in white papers available on the ABA website[9–14] and in peer-reviewed publications[15–19].

The ABA is a genome-wide survey, for which an industrialized 'one size fits all' protocol was used (FIG. 1). For the most part, each gene was assayed throughout the entire left hemisphere in a single brain that was sectioned in the sagittal plane. Approximately 4,000 genes were also analysed throughout a second brain that was coronally sectioned. ISH probes were designed to confer maximal signal detection and specificity and were, on average, 800

nucleotides in length. Nevertheless, exquisitely specific probes could not be designed for all transcripts. For some gene families, such as the zinc finger proteins, sufficiently unique probes could not be generated, and the data for such genes are therefore not in the Atlas. In other cases, a probe may cross-hybridize to another gene. Such cases have been brought to the attention of the ABA through internal data review as well as community feedback via the website and at conferences. For example, a small 120-nucleotide region of the oxytocin gene contained in the original probe design was identical to a region in the vasopressin gene, resulting in cross-hybridization. When this issue was recognized, a new probe was designed and new ISH data were obtained. Finally, probes could detect multiple splice forms of the target gene, such as for muscarinic acetylcholine receptor 2 (REF. 20), and should therefore be assumed to report one or more alternatively spliced transcripts. To address these issues, sequence information for each probe is provided in the ABA so that users can judge specificity for themselves based on currently available gene annotations.

To provide true cellular resolution, a non-isotopic colorimetric ISH assay was used to generate the data, yielding a blue/purple precipitate in the somata of cells expressing the gene of interest. By contrast, isotopic methods deposit silver grains in the proximity of expressing cells. The colorimetric assay is nonlinear, as signal detection requires amplification at several steps. Although it has been shown for specific genes to be as sensitive as radiometric approaches[21–24], it is only semi-quantitative in its output, allowing some discrimination of low, medium or high levels of expression[16,19]. For the ABA, signal amplification and detection times were optimized to detect low-level expressors; therefore the ISH signal can be saturated for more abundantly expressed genes.

Despite efforts to use well-designed probes and reveal low-level transcripts, false negatives are to be expected. No single protocol or methodology is likely to detect all transcripts in all brain structures. Furthermore, the informatics computations, which match manually derived positive or negative designations[15], were designed to minimize false positives and so might miss some regions of expression. Nevertheless, the ABA reveals a greater number of expressed genes than previously observed. Approximately 80% of all genes exhibit some expression in the ABA[15], in contrast to the 55% detected in the brain using

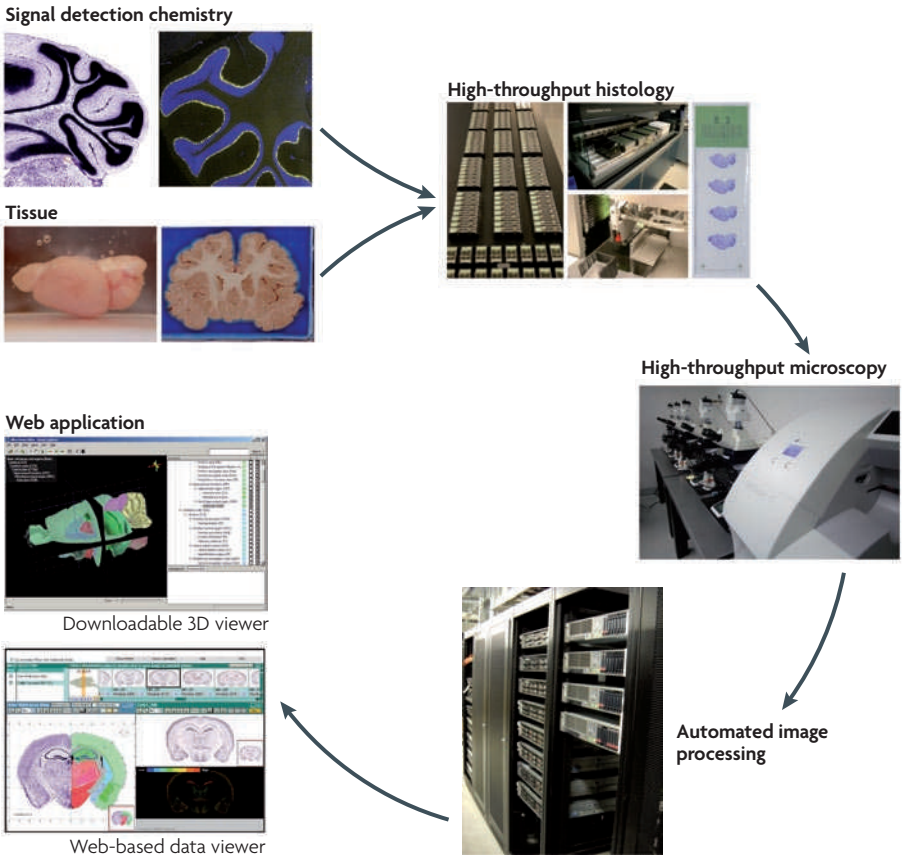


Figure 1 | **The Allen Institute technology platform.** The technology platform begins with two inputs, tissue and signal detection chemistry. Tissue includes brain tissue from mouse, rat, primate and human as well as spinal cord and tumour tissue. Signal detection chemistry indicates the detection of RNA using single digoxigenin riboprobes (colorimetric *in situ* hybridization (ISH)) or two fluorescent riboprobes (double fluorescent ISH) as well as the detection of protein expression by antibodies with immunohistochemistry or microscopic structures with histological stains, such as Nissl. The high-throughput histology platform combines the tissue and signal detection chemistry and consists of cryosectioning, post-cryosection processing (fixation, acetylation and dehydration), ISH using a modified version of the Tecan GenePaint system, and automated coverslipping. To maximize throughput with the adult mouse brain, four sections are placed on each slide. The signal from the riboprobes is captured by either brightfield or fluorescent high-throughput microscopy with either an image capture system or an Aperio ScanScope scanning system. After microscopy, the scanned images are automatically processed through the informatics data pipeline and manually inspected for processing artefacts as part of quality control procedures. The informatics data pipeline consists of preprocessing modules as well as registration and expression detection modules. A storage area network provides online access to all image data and a 148 CPU Linux cluster facilitates high-throughput data processing. The data is publicly displayed via Web-based data viewers and downloadable through various freely available tools, such as the Brain Explorer 3D viewer.

microarrays[25]. Future experiments across the research community, using different protocols or methodologies, will ultimately reveal which of the remaining 20% of genes are expressed in the brain.

Some voices in the neuroscience community have contended that a significant proportion of the ABA data is 'wrong', and some users have documented discrepancies with other studies[26]. To ensure with high confidence that the ABA data is representative for the specific probe design and experimental conditions described, the 1,000 most frequently accessed ABA genes have been compared with other publicly available data (including an extensive scientific literature review for the top 100 genes). This confirmed that ABA expression patterns are similar to those described by others[27] (Supplementary information S1 (box)). In the few cases in which a discrepancy was found, it could often be explained by differences in probe design, cross-species comparison or differences in the platform.

There is also an active process of 'probe discrepancy reporting', in which any feedback through the website opens an

investigation. In most cases, such investigations result in probe redesign and re-running data for a gene. If the new data better represent the pattern of expression in comparison with external results or signal-to-noise on the ABA platform, it is added to the ABA site and the old data are flagged, leaving them accessible in case they have been used, linked to or cited previously.

Any data platform, particularly a high-throughput platform, has a certain rate of error. There were data quality issues early on, such as dissection damage, high background relative to ISH signal, and suboptimal imaging results, reflecting the challenges of converting bench-top laboratory techniques to industrialized, high-throughput processes. As is common for any experimental process in any laboratory, significant improvements in practices and data quality were made throughout the course of the project and beyond. Overall, as quality improved, ISH for approximately 15% of the genes was repeated with redesigned probes or other process improvements.

How the atlas is used

Website traffic statistics document the widespread and extensive use of the ABA, with over 10,000 distinct users accessing the ABA each month on average, and approximately 1,000 users visiting on an average workday. Users from all types of research organizations — academic, governmental, pharmaceutical, biotechnological and non-profit — and from across the globe frequent the ABA website. Mining the ABA Web logs reveals that the entire genome-wide data set is being used. Even before bulk programmatic access to the ABA was enabled with the release of the API (see below), all genes had been accessed (Supplementary information S2 (figure)). PANTHER (protein analysis through evolutionary relationships) analysis[28] of the 1,000 most frequently accessed genes reveals that ion channels and G-protein coupled receptors are the two most prevalent molecular function categories represented, with approximately 90 genes each (Supplementary information S1 (box)).

Site navigation patterns, analysed using Pathalizer, differ between academic and industry users. Industry users tend to access the data by navigating directly to specific gene and image series pages, performing wide data mining and browsing using the ABA's integrated mining tools, and using the sagittal reference atlas. By contrast, academic users usually perform direct gene searches, with less general data mining and browsing, and use the coronal reference atlas more than the sagittal one.

Although Web traffic logs reveal navigation patterns and operational trends in how users interact with the ABA, the openness of the data access and the anonymity of users limits the ability to determine what the thousands of visitors are doing with this resource. Understanding its usage therefore relies heavily on user feedback, including word-of-mouth feedback, written comments and questions submitted via the website, literature citations of the primary atlas publication[15] and citations of the website itself. This feedback indicates extensive 'behind-the-scenes' use of the ABA as a tool to facilitate and direct research, as well as to support grant applications, even if the actual ABA data is not incorporated into resulting published work. The ABA helps investigators focus and select promising avenues for exploration, design transgenic mouse lines, generate and refine hypotheses, and check their own data.

Corroborating results and promoting discovery. There were over 300 citations of the primary ABA publication[15] in the literature as of August 2009, reflecting a broad range of applications throughout neuroscience. The literature shows ABA use supporting not only mouse studies but also those from other species, including rat[29–33], human[30,34,35], *Caenorhabditis elegans*[36] and *Drosophila melanogaster*[37]. Often, these

Box 1 | **Published applications of the ABA dataset in the scientific community**

Medium- and large-scale applications

- Registration and analysis of Allen Brain Atlas (ABA) data[55].
- Comparison of expression patterns of adhesion G protein-coupled receptors in mouse and rat[56].
- Examination of expression patterns of genes identified as regionally enriched by serial analysis of gene expression[57].
- Review of 42 genes involved with feeding neuroregulators[58].
- Study of neocortical malformations in inbred mice[59].
- Data verification for a transcriptome database for astrocytes, neurons and oligodendrocytes[39].
- Prediction of the cell type distribution of age-related gene expression changes in the human brain[60].
- Identification of anterior cingulate cortex-enriched genes to study the molecular mechanism of pain affection[61].
- Verification of mesencephalic dopaminergic system expression[46].
- Expression analysis of genes expressed in bipolar cells in the retina[62].
- Examination of selenoprotein expression in 159 brain regions[63].

Small-scale applications

- Study of addiction (including studies of prevalence of alcohol addiction in mice, cocaine addiction, protein kinases and addiction, neurexin 3 polymorphisms and alcohol dependence, acute alcohol tolerance, and genome-wide association studies of addiction)[40,41,64–67].
- Study of the role of nicotinic acetylcholine receptors in the regulation of synaptic transmission and plasticity[68].
- Alzheimer's disease research (the role of Wnt signalling, the link between *GAB2* alleles and Alzheimer's disease risk and the role of retromers)[69–72].
- Examination of genes that protect against α-synuclein misfolding in a *Caenorhabditis elegans* Parkinson's disease model[73].
- Gene expression analysis of the brain tissue of a mouse model of phenylketonuria[74].
- Identification and characterization of phosphoserine as an enhancer of neurogenesis[75].
- Expression analysis of MECP2 (REF. 76).
- Wnt gene family expression analysis[77].
- Colocalization of MTAP2 and TREK channels in the brain[78].
- Analysis of PAX6 expression by astrocytes[79].
- Verification of the IGFBP5 anterior–posterior expression gradient in the cerebral cortex[80].
- Verification of miRNA target genes[81].
- Global expression analysis shows that ephrin B3, but not ephrin B2, is expressed in the brainstem[82].
- Analysis of major histocompatibility complex class 1 expression in the cerebellum[83].
- Review of myostatin's cellular actions, evolution, functional divergence and novel actions[84].
- Annotation of MGLUR4 (also known as GRM4) expression in the subventricular zone[75].

studies use the ABA either to confirm gene expression patterns seen in non-mouse species or for cross-species comparison. One widespread application of the ABA is to validate other data modalities. These include reporter expression in transgenic mouse lines[38], microarray data[31,39,40], real-time quantitative PCR data[41], data on protein abundance and expression[32,42–44], data linking protein expression with classical quantitative trait loci[45], data on neurotransmitter systems obtained using cDNA-based differential display and microarrays[46], and results associating human functional MRI with genetic association studies[47].

The ABA has also been used to foster new scientific discovery. Such uses take several forms, including the identification of new genes associated with specific brain regions or new areas of gene activity and pathway discovery (for example, by elucidating overlap between molecular members of known pathways). Large- and small-scale applications of the data set have contributed to studies ranging from gene or gene family analyses to studies of cell types, brain organization, behaviour and disease (BOXES 1,2).

ABA riboprobe use. At the time that the ABA project was launched, a complete clone collection of known mouse genes was lacking. To maximize genomic coverage, template material for the production of the riboprobes used to generate the atlas was obtained from multiple sources, including the NIH Mammalian Gene Collection[48], the RIKEN FANTOM3 clone collection[49], mouse brain cDNA and mouse genomic DNA. The resulting riboprobe collection has since been requested for other projects in the community. Substantial portions of the ABA probe template collection have been provided to other large-scale efforts, including EURExpress and EMBLYS (embryonic mouse in bioinformative lyceum system)[50].

Educational use. The ABA is being used to enhance neuroscience education, particularly at the undergraduate level, where it offers dynamic tools for teaching neuroanatomy and opportunities for *in silico* laboratory exercises. The vast collection of quality histology and ISH data provide opportunities for engaging students in investigative data analysis and demonstrating cytoarchitecture and cellular diversity[51]. Additionally, the interactive Brain Explorer three-dimensional viewer has been used to present the gross structural organization of the brain and as an entry point for exploring the brain's molecular organization[52].

Box 2 | **Understanding feeding circuitry with the ABA**

It is increasingly recognized that complex behaviours are likely to be governed by the combined activities of multiple brain structures and genes, and that candidate-gene or structure-based approaches to understanding such behaviours have limitations. In a recent study, the Allen Brain Atlas (ABA) was used to analyse the network of anatomic and molecular elements underlying feeding control in order to elucidate a more complete picture of the neural basis of consummatory behaviour[58].

Previous studies of feeding behaviour had generally focused on single brain structures or single genes, peptides or receptors. Although important, these studies have led to narrow models of central control of feeding behaviour, assigning specific genes specific roles, and similarly viewing specific brain structures as centres governing particular aspects of consumption. This study focused on 8 brain structures and 42 genes known to be involved in consummatory behaviour. Gene expression patterns were analysed and scored for expression level and density in each structure using the Brain Explorer 3D viewer, in which multiple genes were viewed simultaneously to help identify co-expressors, and corresponding two-dimensional images of the original *in situ* hybridization data. Expression networks of both the orexigenic and the anorexigenic genes in the central feeding circuitry were derived from this analysis.

The authors concluded that the expression distribution of the 42 genes strongly supports a widespread central network model for the neural regulation of consummatory behaviour. Each analysed structure contains expressed genes that are associated with a range of feeding functions, and almost every gene is expressed in multiple structures, probably affecting multiple circuitry components.

This example shows that the ABA provides a tool for expanding beyond an incomplete picture of neural control of behaviour derived from candidate approaches and for uncovering complex combinatorial neural machinery. By exposing these systems, the ABA may help to form new, more effective therapeutic strategies that target multiple facets of a system simultaneously.

Programmatic access via the API. Over time, more useful avenues for public access to the ABA have been engineered. The addition of machine-accessible Extensible Markup Language (XML) files and an API allows users to access the data underlying the ABA, including three-dimensional coordinates and informatically derived expression values, download it into their own computational environment and create tools, new Web or software applications that combine ABA data with other resources (mashups), or other applications. Already, several such uses have appeared in the public domain. Researchers at Janelia Farm have created ALLENMINER, a public tool to identify genes expressed with certain patterns in the ABA[53]. UCLA's Laboratory of Neuro Imaging is creating a uniform Web services API for their Mouse BIRN Atlasing Toolkit architecture and, with assistance from the NeuroCommons project, is using the API along with GENSAT, GeneNetwork and others to support data queries. Science Commons, as part of the NeuroCommons project, has created a demonstration mashup that allows viewing of ABA data in Google Maps, and the brainmaps.org Ajax viewer and API can view ABA data.

Unexplored territory

At its heart, the ABA is a genomics project. It is the first comprehensive investigation across the genome of where genes are expressed in a complex biological structure. However, the full potential of the data set for research in the genomics community has not been realized. This is partly due to the initial structuring and presentation of the data as indexed sets of images, an unfamiliar data modality to genomics researchers. Indeed, gene-by-gene viewing of two-dimensional ISH image data through browser pages, thumbnail views and higher-resolution views accounts for 98% of all site traffic. Additional obstacles to adoption included the focus on the brain and the coarse resolution of the initial informatics-based data mining tools, which mapped the expression data to only a few brain structures, in contrast to the later higher-resolution voxel-based mapping.

To create a framework for genomic-style data mining of the ABA, the original two-dimensional tissue sections have been mapped into a common three-dimensional anatomic framework. The data have been abstracted into a matrix of approximately 20,000 genes × 50,000 coordinates in a volumetric standardized space, providing the foundation for innovative mining approaches that integrate the genomic and anatomical dimensions of the ABA. The resulting data matrix is openly available through the API and is ready to be incorporated into various genome analysis strategies, from comparative evolutionary genomics to promoter mapping.

Box 3 | **Advanced features and data mining tools**

NeuroBlast
Derived from the concept behind the Basic Local Alignment Search Tool (BLAST)[85] that is used to identify genes with sequences similar to a gene of interest, NeuroBlast identifies genes with similar patterns of expression in the brain[86]. Starting with a gene of interest, users can generate a ranked list of genes with similar three-dimensional (3D) expression patterns. For example, seeding NeuroBlast with calbindin 1 returns other genes with similarly enriched expression in Purkinje cells. NeuroBlast covers 20 different anatomical regions, allowing users to conduct brain-wide searches or to tailor searches to a region of interest.

Allen Reference Atlas (ARA)
Developed to align specifically with the Allen Brain Atlas (ABA) *in situ* hybridization data, the ARA[87] comprises a full-colour, high-resolution, Web-based digital neuroanatomical atlas accompanied by a systematic, hierarchically organized taxonomy of mouse brain structures. The ARA includes coronal and sagittal plates and is fully integrated into the ABA, allowing a direct comparison of gene expression patterns to neuroanatomical structures. The ARA also provides a standard neuroanatomical ontology and 3D coordinate framework for determining structural annotation and enabling anatomical search and navigation of the gene expression database.

Brain Explorer 3D viewer
Brain Explorer is an interactive tool for visualizing mouse brain anatomy and ABA gene expression data in the framework of the ARA[17]. Users view an interactive, 3D rendering of the mouse brain and can rotate views, virtually slice in three planes, turn individual brain structures on and off, and view expression patterns of one or more genes. NeuroBlast searches and direct access to the raw ABA data on which the 3D renderings are based are available.

Anatomic Gene Expression Atlas (AGEA)
Based on computed spatial correlations across expression data for thousands of genes, the AGEA is an interactive 3D atlas that reveals brain organization based on the spatial organization of the transcriptome[18]. The AGEA offers users tools for exploring neuroanatomical relationships and boundaries and the molecular organization of the brain: 3D gene expression-based correlation maps, hierarchical transcriptome-based parcellations of the brain and a search function to retrieve a list of specific genes showing enriched expression within local correlated domains.

Fine Structure Annotation (FSA) gene list
Curated lists of genes that are enriched in each of more than 70 specific neuroanatomical structures are available. These lists are sets of approximately 50 genes with specific expression patterns in the structure of interest. For example, in the hippocampus, FSA gene lists are available for the dentate gyrus, field CA1 pyramidal layer, field CA2 pyramidal layer, field CA3 pyramidal layer, the subiculum and the entorhinal area. For some structures, users can also access FSA reports that offer deeper annotation of the gene expression data, including detailed descriptions of each region, the characteristics of the genes that are selective for the region and correlation tables showing the genetic relationship of the region to the rest of the brain.

Over time, the data have been annotated more deeply, and powerful mining and visualization tools that reach across the data set as a whole and capitalize on the robust spatial mapping of genomic information have been added (BOX 3). These features enable new searches and analyses that allow users to access and browse greater quantities of relevant data more quickly. Nevertheless, perhaps because they were initially difficult to find, not well publicized and secondary to the data itself, such tools are currently underutilized and account for only 2% of site traffic.

Although the lack of user registration requirements for the ABA supports the goal of open access, the resulting anonymity of users poses a challenge for informing them of enhancements. To better communicate new tools and content to users and make them easier to find, the user interface design was improved, and a new ABA web portal providing access to all datasets and news about the latest additions was launched in July 2008. Additional opportunities to improve communication to users continue to be explored.

Resources for the future

After completing the ABA, the Allen Institute embarked on a series of additional large-scale projects, all of which are yielding free, publicly available online resources (TIMELINE) accessible through the ABA portal. Completed projects include a large-scale study on sleep deprivation and diurnal variation of gene expression in the adult mouse brain; a genome-wide atlas of the mouse spinal cord that includes ISH, histology and reference atlases in early postnatal and adult mice; a mouse diversity study, with ISH data for 48 genes encoding drug targets in males and females, and across males in 7 strains of mice; a two-part human cortex study comprising a survey of over 1,000 genes in multiple neocortical regions and a 60-gene ISH study in the dorsolateral prefrontal cortex across multiple control and schizophrenia subjects.

Ongoing projects include an atlas of the developing mouse brain comprising ISH data for approximately 2,000 genes and reference atlases across multiple stages of prenatal development, postnatal development and aging; an atlas of the developing postnatal Rhesus macaque brain (sponsored by the NIH Blueprint for Neuroscience Research); and a multi-modal atlas of the whole human brain, the first phase of which will contain structural MRI, histology and microarray-based gene expression data for 1,000 distinct anatomical areas from each of two human brains. Later phases of the human brain atlas are expected to incorporate high-resolution ISH data across several brain regions.

With the growing availability of public neuroscience resources online, the next major challenge will be to effectively integrate them in order to maximize their accessibility, usability and value to researchers. The Allen Institute is actively exploring ways to enable users to more easily discover related data and navigate across all of its online resources. In addition, the recently developed ABA API sets the stage for community-driven integration of the ABA with other resources. Mashups using the API have already begun to appear publicly, as described above, and it will be interesting to see what additional innovations emerge. More broadly, various community efforts have been initiated to develop standards and strategies to achieve better crosstalk and data integration across neuroscience resources. These include task forces in digital brain atlasing and in neural structure ontologies established by the International Neuroinformatics Coordinating Facility and additional efforts through the Neuroscience Information Framework, an NIH Blueprint for Neuroscience Research initiative.

Conclusions

Large-scale neuroscience is in its infancy. The creation of high-value resources that serve the broad neuroscience community requires tackling not only challenges faced by classic genomics projects but also additional challenges particular to the field of neuroscience: the multi-modal and multi-resolution nature of the data — which ranges from whole-brain imaging and gross anatomy to cellular-resolution

Box 4 | **Venture philanthropy for large-scale atlasing initiatives**

In contrast to the prevailing individual-investigator model, the Allen Institute undertakes large, enterprise-wide projects like the Allen Brain Atlas (ABA), relying on integrated efforts and expertise across its multidisciplinary staff. Each project is executed following an industrial model, with full accountability to scope, budget, schedule and productivity benchmarks defined at the outset. These projects have a significant emphasis on an end-product deliverable and result in community resources intended to help advance neuroscience research programmes worldwide.

Despite its unique operational model, like other non-profit medical research organizations the Allen Institute currently supports its projects through a mixture of private and public funds. Paul Allen's initial seed funding launched the Allen Institute and funded its inaugural ABA project, demonstrating the feasibility of such large-scale atlasing efforts and their value to the broad neuroscience research community. From the beginning, it was expected that the Allen Institute would raise funds from multiple sources to support future initiatives. In line with that goal, the second major atlas project to come online, the Allen Spinal Cord Atlas, was made possible by a diverse consortium of funders, including foundations, disease organizations, corporations and private donors. Other projects have also received funding from public sources, such as the US National Institutes of Health. Following a project-focused funding model, these funds, regardless of their source, are converted to tangible deliverables that are accessible to funders and the community at large.

histology and physiology — and the complex, three-dimensional nature of the brain. Neuroanatomical atlases provide the most logical framework for integrating various data modalities and organizing and presenting the data for the broad, multidisciplinary neuroscience community. As a consequence, the creation of such resources involves computational resources and expertise not traditionally found in neuroscience laboratories outside of those working on brain imaging: high-throughput image management, sophisticated image databasing, software development and informatics.

The completion of the ABA and its incorporation into the workflows of wide-ranging neuroscientists has shown that large-scale science can work in the neurosciences. The ultimate value of such resources is dependent on the extensive community of basic researchers that they support, catalysing progress and efficiency in large and small ways across many subdisciplines. Built on an industrial operational model (BOX 4) and fuelling basic research progress, the Allen Institute provides a unique framework for generating, providing and maintaining such resources that are hard to create in academia, although the genome sequencing centres offer valuable templates. Calls for such centres and large-scale science[54] are certainly well placed as they encourage progress. However, such approaches are highly complex and not simply born from a cohort of several research technicians and a project manager. They require broad expertise and skills across many different disciplines, careful orchestration, a culture of teamwork and a strong 'customer' focus.

Allan R. Jones, Caroline C. Overly and Susan M. Sunkin are at the Allen Institute for Brain Science, 551 North 34th Street Seattle, Washington 98103, USA.
Correspondence to A.R.J.
e-mail: AllanJ@alleninstitute.org

doi:10.1038/nrn2722
Published online 14 October 2009

1. Tessier-Lavigne, M. & Stryer, L. Setting priorities for molecular neuroanatomy in the postgenomic era. *NIH NIDA* [online], http://www.drugabuse.gov/MeetSum/Postgenomic.html (2002).
2. Heintz, N. Gene expression nervous system atlas (GENSAT). *Nature Neurosci.* **7**, 483 (2004).
3. Geschwind, D. GENSAT: a genomic resource for neuroscience research. *Lancet Neurol.* **3**, 82 (2004).
4. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
5. Tecan. The GenePaint System. *Mol. Biotechnol.* **25**, 103–104 (2003).
6. Mamounas, L., Gubitz, A. K. & Talley, N. NIH blueprint for neuroscience research: GENSAT. *NIH* [online], http://neuroscienceblueprint.nih.gov/factsheet/gensat.pdf (2008).
7. Burris, J., Cook-Deegan, R. & Alberts, B. The Human Genome Project after a decade: policy issues. *Nature Genet.* **20**, 333–335 (1998).
8. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286–290 (2003).
9. Allen Institute for Brain Science. ISH platform controls. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/ISH_Platform_Controls.pdf (2006).
10. Allen Institute for Brain Science. Cross platform validation. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/Cross_Platform_Validation.pdf (2006).
11. Allen Institute for Brain Science. Informatics data processing. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/InformaticsDataProcessing.pdf (2006).
12. Allen Institute for Brain Science. Data production processes. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/ABADataProductionProcesses.pdf (2006).
13. Allen Institute for Brain Science. NeuroBlast: user guide. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/NeuroBlast.pdf (2007).
14. Allen Institute for Brain Science. AGEA user guide. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/AGEA.pdf (2007).
15. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
16. Ng, L. L. *et al.* Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE/ACM Trans Comput. Biol. Bioinform.* **4**, 382–393 (2007).
17. Lau, C. *et al.* Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics* **9**, 153–163 (2008).
18. Ng, L. *et al.* An anatomic gene expression atlas of the adult mouse brain. *Nature Neurosci.* **12**, 356–362 (2009).
19. Lee, C. K. *et al.* Quantitative methods for genome-scale analysis of *in situ* hybridization and correlation with microarray data. *Genome Biol.* **9**, R23 (2008).
20. Gosso, F. M. *et al.* Exploring the functional role of the CHRM2 gene in human cognition: results from a dense genotyping and brain expression study. *BMC Med. Genet.* **8**, 66 (2007).
21. Hougaard, D. M., Hansen, H. & Larsson, L. I. Non-radioactive *in situ* hybridization for mRNA with emphasis on the use of oligodeoxynucleotide probes. *Histochem. Cell Biol.* **108**, 335–344 (1997).
22. Higo, N., Oishi, T., Yamashita, A., Matsuda, K. & Hayashi, M. Quantitative non-radioactive *in situ* hybridization study of GAP-43 and SCG10 mRNAs in the cerebral cortex of adult and infant macaque monkeys. *Cereb. Cortex* **9**, 317–331 (1999).
23. Bello, M. *et al.* Hybrid segmentation framework for tissue images containing gene expression data. *Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* **8**, 254–261 (2005).
24. Carson, J. P. *et al.* A digital atlas to characterize the mouse brain transcriptome. *PLoS Comput. Biol.* **1**, e41 (2005).
25. Sandberg, R. *et al.* Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).
26. Heimel, J. A., Hermans, J. M., Sommeijer, J. P. & Levelt, C. N. Genetic control of experience-dependent plasticity in the visual cortex. *Genes Brain Behav.* **7**, 915–923 (2008).
27. Allen Institute for Brain Science. Comparison of the top 1000 genes. *Allen Brain Atlas* [online], http://mouse.brain-map.org/pdf/Top1000GenesAnalysis.pdf (2007).
28. Mi, H., Guo, N., Kejariwal, A. & Thomas, P. D. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.* **35**, D1247–D1252 (2007).
29. McHugh, P. C. *et al.* Downregulation of *Ccnd1* and *Hes6* in rat hippocampus after chronic exposure to the antidepressant paroxetine. *Acta Neuropsychiatrica* **20**, 307–313 (2008).
30. Ball-Rosen, C. *et al.* Identification of histidine-rich glycoprotein, a potential autoantigen, in human and rat brain preparations. *Ann. NY Acad. Sci.* **1109**, 473–483 (2007).
31. Greene, J. G., Borges, K. & Dingledine, R. Quantitative transcriptional neuroanatomy of the rat hippocampus: evidence for wide-ranging, pathway-specific heterogeneity among three principal cell layers. *Hippocampus* **19**, 253–264 (2009).
32. Sakakibara, S. *et al.* Developmental and spatial expression pattern of alpha-taxilin in the rat central nervous system. *J. Comp. Neurol.* **511**, 65–80 (2008).
33. Benoist, M. *et al.* Distribution of zinedin in the rat brain. *J. Neurochem.* **106**, 969–977 (2008).
34. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).
35. Ernst, C. *et al.* Confirmation of region-specific patterns of gene expression in the human brain. *Neurogenetics* **8**, 219–224 (2007).
36. Von Stetina, S. E. *et al.* Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the *C. elegans* nervous system. *Genome Biol.* **8**, R135 (2007).
37. Mecklenburg, K. L. *Drosophila* retinophilin contains MORN repeats and is conserved in humans. *Mol. Genet. Genomics* **277**, 481–489 (2007).
38. Chizhikov, V. V. *et al.* Cilia proteins control cerebellar morphogenesis by promoting expansion of the granule progenitor pool. *J. Neurosci.* **27**, 9780–9789 (2007).
39. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
40. Tabakoff, B. *et al.* The genomic determinants of alcohol preference in mice. *Mamm. Genome* **19**, 352–365 (2008).
41. Kelai, S. *et al.* Nrxn3 upregulation in the globus pallidus of mice developing cocaine addiction. *Neuroreport* **19**, 751–755 (2008).

42. Andrade, N. *et al.* ApoER2/VLDL receptor and Dab1 in the rostral migratory stream function in postnatal neuronal migration independently of Reelin. *Proc. Natl Acad. Sci. USA* **104**, 8508–8513 (2007).
43. Satoh, J. *et al.* Protein microarray analysis identifies human cellular prion protein interactors. *Neuropathol. Appl. Neurobiol.* **35**, 16–35 (2009).
44. Lim, J. *et al.* Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* **452**, 713–718 (2008).
45. Mozhui, K. *et al.* Dissection of a QTL hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. *PLoS Genet.* **4**, e1000260 (2008).
46. Alavian, K. N. & Simon, H. H. Linkage of cDNA expression profiles of mesencephalic dopaminergic neurons to a genome-wide *in situ* hybridization database. *Mol. Neurodegener.* **4**, 6 (2009).
47. Papassotiropoulos, A. *et al.* Common Kibra alleles are associated with human memory performance. *Science* **314**, 475–478 (2006).
48. Gerhard, D. S. *et al.* The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* **14**, 2121–2127 (2004).
49. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
50. Sunkin, S. M. & Hohmann, J. G. Insights from spatially mapped gene expression in the mouse brain. *Hum. Mol. Genet.* **16**, R209–R219 (2007).
51. Ramos, R. L., Smith, P. T. & Brumberg, J. C. Novel *in silico* method for teaching cytoarchitecture, cellular diversity, and gene expression in the mammalian brain. *Journal of Undergraduate Neuroscience Education* **6**, A8–A13 (2007).
52. Jenks, B. G. A self-study tutorial using the Allen Brain Explorer and Brain Atlas to teach concepts of mammalian neuroanatomy and brain function. *Journal of Undergraduate Neuroscience Education* **8**, A21–A25 (2009).
53. Davis, F. P. & Eddy, S. R. A tool for identification of genes expressed in patterns of interest using the Allen Brain Atlas. *Bioinformatics* **25**, 1647–1654 (2009).
54. Lichtman, J. W. & Sanes, J. R. Ome sweet ome: what can the genome tell us about the connectome? *Curr. Opin. Neurobiol.* **18**, 346–353 (2008).
55. Jagalur, M., Pal, C., Learned-Miller, E., Zoeller, R. T. & Kulp, D. Analyzing *in situ* gene expression in the mouse brain with image registration, feature extraction and block clustering. *BMC Bioinformatics* **8** (Suppl. 10), S5 (2007).
56. Haitina, T. *et al.* Expression profile of the entire family of Adhesion G protein-coupled receptors in mouse and rat. *BMC Neurosci.* **9**, 43 (2008).
57. D'Souza, C. A. *et al.* Identification of a set of genes showing regionally enriched expression in the mouse brain. *BMC Neurosci.* **9**, 66 (2008).
58. Olszewski, P. K., Cedernaes, J., Olsson, F., Levine, A. S. & Schioth, H. B. Analysis of the network of feeding neuroregulators using the Allen Brain Atlas. *Neurosci. Biobehav. Rev.* **32**, 945–956 (2008).
59. Ramos, R. L. *et al.* Cytoarchitecture and transcriptional profiles of neocortical malformations in inbred mice. *Cereb. Cortex* **18**, 2614–2628 (2008).
60. Loerch, P. M. *et al.* Evolution of the aging brain transcriptome and synaptic regulation. *PLoS ONE* **3**, e3329 (2008).
61. Sun, Y. G. *et al.* Involvement of P311 in the affective, but not in the sensory component of pain. *Mol. Pain* **4**, 23 (2008).
62. Kim, D. S. *et al.* Identification of molecular markers of bipolar cells in the murine retina. *J. Comp. Neurol.* **507**, 1795–1810 (2008).
63. Zhang, Y. *et al.* Comparative analysis of selenocysteine machinery and selenoproteome gene expression in mouse brain identifies neurons as key functional sites of selenium in mammals. *J. Biol. Chem.* **283**, 2427–2438 (2008).
64. Uhl, G. R. *et al.* Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Ann. NY Acad. Sci.* **1141**, 318–381 (2008).
65. Lee, A. M. & Messing, R. O. Protein kinases and addiction. *Ann. NY Acad. Sci.* **1141**, 22–57 (2008).
66. Hishimoto, A. *et al.* Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. *Hum. Mol. Genet.* **16**, 2880–2891 (2007).
67. Hu, W. *et al.* Genomic insights into acute alcohol tolerance. *J. Pharmacol. Exp. Ther.* **326**, 792–800 (2008).
68. McKay, B. E., Placzek, A. N. & Dani, J. A. Regulation of synaptic transmission and plasticity by neuronal nicotinic acetylcholine receptors. *Biochem. Pharmacol.* **74**, 1120–1133 (2007).
69. Inestrosa, N. C. & Toledo, E. M. The role of Wnt signaling in neuronal dysfunction in Alzheimer's disease. *Mol. Neurodegener.* **3**, 9 (2008).
70. Reiman, E. M. *et al.* GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* **54**, 713–720 (2007).
71. Muhammad, A. *et al.* Retromer deficiency observed in Alzheimer's disease causes hippocampal dysfunction, neurodegeneration, and Aβ accumulation. *Proc. Natl Acad. Sci. USA* **105**, 7327–7332 (2008).
72. Small, S. A. Retromer sorting: a pathogenic pathway in late-onset Alzheimer disease. *Arch. Neurol.* **65**, 323–328 (2008).
73. Hamamichi, S. *et al.* Hypothesis-based RNAi screening identifies neuroprotective genes in a Parkinson's disease model. *Proc. Natl Acad. Sci. USA* **105**, 728–733 (2008).
74. Park, J. W., Park, E. S., Choi, E. N., Park, H. Y. & Jung, S. C. Altered brain gene expression profiles associated with the pathogenesis of phenylketonuria in a mouse model. *Clin. Chim. Acta* **401**, 90–99 (2009).
75. Saxe, J. P. *et al.* A phenotypic small-molecule screen identifies an orphan ligand-receptor pair that regulates neural stem cell differentiation. *Chem. Biol.* **14**, 1019–1030 (2007).
76. Belichenko, N. P., Belichenko, P. V., Li, H. H., Mobley, W. C. & Francke, U. Comparative study of brain morphology in Mecp2 mutant mouse models of Rett syndrome. *J. Comp. Neurol.* **508**, 184–195 (2008).
77. Liu, C., Wang, Y., Smallwood, P. M. & Nathans, J. An essential role for Frizzled5 in neuronal survival in the parafascicular nucleus of the thalamus. *J. Neurosci.* **28**, 5641–5653 (2008).
78. Sandoz, G. *et al.* Mtap2 is a constituent of the protein network that regulates twik-related K^+ channel expression and trafficking. *J. Neurosci.* **28**, 8545–8552 (2008).
79. Sakurai, K. & Osumi, N. The neurogenesis-controlling factor, Pax6, inhibits proliferation and promotes maturation in murine astrocytes. *J. Neurosci.* **28**, 4604–4612 (2008).
80. Gatchel, J. R. *et al.* The insulin-like growth factor pathway is altered in spinocerebellar ataxia type 1 and type 7. *Proc. Natl Acad. Sci. USA* **105**, 1291–1296 (2008).
81. Glazov, E. A., McWilliam, S., Barris, W. C. & Dalrymple, B. P. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.* **25**, 939–948 (2008).
82. Negrete, O. A., Chu, D., Aguilar, H. C. & Lee, B. Single amino acid changes in the Nipah and Hendra virus attachment glycoproteins distinguish ephrinB2 from ephrinB3 usage. *J. Virol.* **81**, 10804–10814 (2007).
83. Letellier, M., Willson, M. L., Gautheron, V., Mariani, J. & Lohof, A. M. Normal adult climbing fiber monoinnervation of cerebellar Purkinje cells in mice lacking MHC class I molecules. *Dev. Neurobiol.* **68**, 997–1006 (2008).
84. Rodgers, B. D. & Garikipati, D. K. Clinical, agricultural, and evolutionary biology of myostatin: a comparative review. *Endocr. Rev.* **29**, 513–534 (2008).
85. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
86. Ng, L. *et al.* NeuroBlast: a 3D spatial homology search tool for gene expression. *BMC Neurosci.* **8**, 11 (2007).
87. Dong, H. W. *The Allen Reference Atlas: A Digital Color Brain Atlas of the C57BL/6J Male Mouse* (Wiley, Hoboken, New Jersey, 2008).

Acknowledgements

This work was sponsored by the Allen Institute for Brain Science. The authors wish to thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement and support.

FURTHER INFORMATION

Allen Brain Atlas portal: http://www.brain-map.org
Allen Developing Mouse Brain Atlas: http://developingmouse.brain-map.org
Allen Institute for Brain Science: http://www.alleninstitute.org
Allen Institute Human Cortex Study: http://humancortex.alleninstitute.org/has
Allen Institute Mouse Diversity Study: http://mousediversity.alleninstitute.org
Allen Institute Sleep Study: http://sleep.alleninstitute.org
Allen Institute Transgenic Mouse Study: http://transgenicmouse.alleninstitute.org
ALLENMINER: http://research.janelia.org/davis/allenminer/
Allen Mouse Brain Atlas (original ABA): http://mouse.brain-map.org/
Allen Spinal Cord Atlas: http://mousespinal.brain-map.org/
Brain Explorer self-guided tutorial for education: http://community.brain-map.org/confluence/display/EDU/Education+Home
EURExpress: http://www.eurexpress.org/ee
GENSAT: http://www.gensat.org
International Neuroinformatics Coordinating Facility: http://www.incf.org
Mouse BIRN Atlasing Toolkit: http://cms.loni.ucla.edu/MBAT/index.aspx
Mutant Mouse Regional Resource Centers (MMRRC): http://www.mmrrc.org
NeuroCommons project: http://sciencecommons.org/projects/data/details
Neuroscience Information Framework: http://nif.nih.gov
Pathalizer: http://pathalizer.sourceforge.net/index.html

SUPPLEMENTARY INFORMATION

See online article: S1 (box) | S2 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Biographies

Allan Jones joined the Allen Institute in 2003 to help start up the organization and is now the Institute's Chief Scientific Officer. Bringing extensive expertise in project leadership and high-throughput genomics operations from prior management positions at Merck and Co. and Rosetta Inpharmatics, he was instrumental in building the Institute's scientific operations from the ground up and successfully driving the Allen Mouse Brain Atlas to completion in 2006. He holds a B.S. degree in biology from Duke University, North Carolina, USA, and a Ph.D. in genetics and developmental biology from Washington University School of Medicine, USA.

Caroline (Kelly) Overly is Research Alliance Manager at the Allen Institute for Brain Science, where she manages interdisciplinary projects across departments and organizations. In addition to leading collaborations, she is currently helping to oversee the development of the Allen Human Brain Atlas, with responsibility for product management, education and user outreach. Overly has a diverse background that spans neuroscience research, science education, business strategy and communications. She earned her A.B. in biology and her Ph.D. in neuroscience from Harvard University, Massachusetts, USA. She has studied and published work on neuronal metabolism, Alzheimer's disease, genomics and neuroscience.

Susan Sunkin started as the Production Manager at the Allen Institute, then moved into data analysis followed by her current position in which she interfaces with external collaborators and performs project management tasks for various projects. Sunkin has worked at Celltech R&D Inc. as the Core Technology Group manager and Genelex Corporation as the Director of Genome Services. She holds a B.A. degree in microbiology from Miami University, Ohio, USA, and a Ph.D. in molecular genetics, biochemistry, and microbiology from the University of Cincinnati, Ohio, USA. She did her postdoctoral training at the University of Cincinnati and the Seattle Biomedical Research Institute.

TOC blurb

000 The Allen Brain Atlas: 5 years and beyond

Allan R. Jones, Caroline C. Overly and Susan M. Sunkin

Five years after the launch of the Allen Brain Atlas, Jones and colleagues describe the challenges faced during the early years of the project, the contributions that it has made to neuroscience research to date and the opportunities for its use in the future.